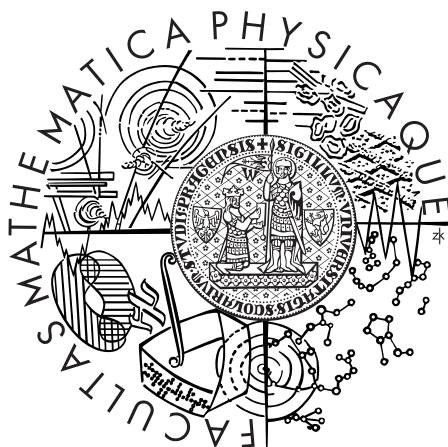


Univerzita Karlova v Praze  
Matematicko-fyzikální fakulta

## BAKALÁŘSKÁ PRÁCE



Nataša Košová

### **Analýza trhu rychloobrátkového zboží**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Soňa Reisnerová, Ph.D.

Studijní program: Matematika, obecná matematika

2009

Rada by som sa poďakovala všetkým, ktorí mi požičali potrebnú literatúru, alebo ma akýmkoľvek spôsobom podporili pri písaní bakalárskej práce. Najmä ďakujem mojej vedúcej RNDr. Soni Reisnerovej, Ph.D., ktorá so mnou podrobne prekonzultovala náplň práce a diskutovala o problémoch pri jej písaní.

Prehlasujem, že som svoju bakalársku prácu napísala samostatne a výhradne s použitím citovaných prameňov. Súhlasím so zapožičiavaním práce a jej zverejňovaním.

V Prahe dňa 7. 8. 2009

Nataša Košová

# Obsah

<b>1</b>	<b>Úvod</b>	<b>5</b>
<b>2</b>	<b>Časová závislosť</b>	<b>7</b>
2.1	Úprava dát . . . . .	7
2.2	Trend . . . . .	9
2.3	Sezónnosť . . . . .	12
2.4	Interpretácia výsledkov . . . . .	14
<b>3</b>	<b>Priestorová závislosť</b>	<b>16</b>
3.1	Úprava dát . . . . .	16
3.2	Popisná štatistika . . . . .	18
3.3	Lineárna regresia . . . . .	22
<b>4</b>	<b>Záver</b>	<b>28</b>
	<b>Literatúra</b>	<b>30</b>
	<b>Príloha 1</b>	<b>31</b>
	<b>Príloha 2</b>	<b>42</b>

Názov práce: Analýza trhu rychloobrátkového zboží

Autor: Nataša Košová

Katedra (ústav): Katedra pravděpodobnosti a matematické statistiky

Vedúci bakalárskej práce: RNDr. Soňa Reisnerová, Ph.D.

e-mail vedúceho: sona.reisnerova@centrum.cz

Abstrakt: V predloženej práci analyzujeme predaj výrobkov v závislosti na čase a mieste nákupu. Študujeme trendovú a sezónnu zložku časového radu, ktorý obsahuje dáta týkajúce sa segmentu sušienok a oblátiek. Ďalej študujeme predajnosť rôznych výrobkov v jednotlivých krajoch Českej republiky a závislosť predaja na životnej úrovni kraja. Nakoniec overujeme hypotézy, či tovar s nízkou cenou je atraktívnejší ako tradičné výrobky, opäť vo vzťahu k rôznym krajom Českej republiky. Práca obsahuje spracovanie a úpravu dodaných dát, popis použitých metód analýz a interpretáciu výsledkov analýzy.

Kľúčové slová: trend, sezónnosť, popisná štatistika, lineárna regresia

Title: Analysis of FMCG market

Author: Nataša Košová

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Soňa Reisnerová, Ph.D.

Supervisor's e-mail address: sona.reisnerova@centrum.cz

Abstract: In the present thesis we analyze product sale in dependence on time and place of purchase. We study trend and seasonal component of time serie, which contains data concerning the section of biscuits and wafers. Further we study the saleability of different products and its dependence on living standard in specific regions of the Czech Republic. Finaly we verify the hypothesis, whether goods with lower prices are more attractive than traditional products, again according to specific regions. The thesis contains preparation, processing and editing of given data, description of used analysis and interpretation of the analysis results.

Keywords: trend, seasonality, descriptive statistics, linear regression

# Kapitola 1

## Úvod

Pod pojmom rýchloobrátkový tovar rozumieme produkty, ktoré sa predávajú rýchlo, za nízke ceny a majú omedzenú dobu trvanlivosti. Aj keď zisk z jednotlivých produktov je relatívne malý, predávajú sa v obrovských množstvách, a preto môže byť celkový zisk veľký. Trh rýchloobrátkového tovaru je nasýteným trhom a aby sa jednotliví výrobcovia na ňom udržali, tak musia svoje výrobky často predávať v akciách, či už s cenovým zvýhodnením alebo zatraktívením v podobe darčeka alebo väčšieho balenia. Medzi rýchloobrátkový tovar zahrňujeme napríklad kozmetiku, lieky, balené potravinárske produkty a nápoje.

V tejto práci sa budeme zaoberať analýzou trhu rýchloobrátkového tovaru, to znamená trhu, na ktorom dochádza k predaju a kúpe už vyššie zmienených produktov. Konkrétne budeme študovať a analyzovať dáta medzinárodných obchodných reťazcov týkajúce sa segmentu sušienok a oblátiek.

Dáta zahrňajú údaje o počte predaných kusov tovaru približne za obdobie štyroch rokov, od augusta 2004 do októbra 2008, v každom kraji Českej republiky. Ku každému z výrobkov ešte máme k dispozícii informáciu o jeho priemernej cene za tieto štyri roky, hmotnosti a zaradení do segmentov. Segmentom pritom myslíme zoskupenie podobných výrobkov od rôznych výrobcov, teda zoskupenie substitútov.

V druhej kapitole budeme pracovať s časovými radmi jednotlivých segmentov. Bude nás zaujímať časová závislosť predaja sušienok. To znamená, že nebudeme rozlišovať, v ktorom kraji sa výrobok predal, ale zoberieme sumu pozorovaných hodnôt pre celú Českú republiku. Stručne popíšeme metódy konštrukcie trendovej a sezónnej zložky v daných časových radoch a nakoniec zhrnieme výsledky, ku ktorým sme pri ich analýze dospeli.

V tretej kapitole sa budeme zaoberať priestorovou závislosťou predaja sušienok. Budeme sledovať predajnosť výrobkov rôznych cien v závislosti na mieste predaja, pre všetky kraje Českej republiky. Nebudeme zohľadňovať čas nákupu, ale zoberieme sumu za celé obdobie štúdie a pre jednoduchosť zanedbáme fakt, že kraje so sebou susedia a teda medzi nimi dochádza k vzájomnej interakcii. Ďalej budeme skúmať závislosť životnej úrovne jednotlivých krajov a cenových hladín sušienok.

V poslednej kapitole, a teda v závere práce, obecnne zhrnieme úspechy a neúspechy uskutočnených analýz v oboch predchádzajúcich kapitolách.

# Kapitola 2

## Časová závislosť

Časový rad je rad hodnôt určitého štatistického znaku usporiadaný podľa časového sledu. Časový rad tiež môžeme chápať ako realizáciu konkrétneho náhodného procesu a pozorované hodnoty ako náhodné veličiny.

V tejto kapitole budeme analyzovať časový rad pozorovaných hodnôt celkovej hmotnosti predaných výrobkov v celej Českej Republike. Predovšetkým nás bude zaujímať trendová a sezónna zložka analyzovaného radu.

### 2.1 Úprava dát

K tomu, aby sme mohli začať analyzovať časový rad, potrebujeme najprv spracovať dáta, ktoré máme k dispozícii. Dáta je potrebné očistiť od očividne chybných údajov. Napríklad, ak je pri výrobku uvedená nulová hmotnosť či cena, odstránime ho z dát a tým zabránime prípadnému skresleniu výsledkov, ktoré pri analýze dostaneme. V ďalšom kroku skumulujeme výrobky v jednotlivých segmentoch, čím si zjednodušíme prácu s dátami a hlavne tým zmenšíme vplyv promócií, ktorý sa odráža v cene a počte predaných výrobkov. Nakoniec si musíme vybrať, ktoré údaje budeme skúmať, teda ktoré údaje o výrobkoch budú tvoriť časové rady.

Jednou z možností nazerania na časový rad je jednoducho pozorovať celkový počet predaných sušienok. Každý segment obsahuje sušienky rôznej gramáže, čo znemožňuje skumulovať údaje o predaných kusoch jednotlivých výrobkov v tejto skupine. Tento problém sa dá odstrániť prevedením všetkých výrobkov na rovnakú hmotnosť a nasledovným prepočítaním počtu predaných kusov. Tým dostaneme pre údaje o predaných kusoch necelé čísla, čo môže pôsobiť trochu mätúco. Ďalšou možnosťou je študovať

celkovú sumu, ktorú kupujúci zaplatili za všetky výrobky v jednotlivých segmentoch. To by znamenalo počty predaných kusov vynásobiť ich priemernou cenou za dobu štúdie. Avšak údaj o priemernej cene by nám dáta mohol nepríjemne skresliť, práve kvôli rôznym znižovaniám cien výrobkov v akciách v priebehu sledovaných rokov. Poslednou uvažovanou možnosťou je možnosť analyzovania údaje o celkovej hmotnosti zakúpených výrobkov. Teda počet predaných kusov u každého výrobku by sme vynásobili jeho hmotnosťou a nakoniec všetky celkové hmotnosti výrobkov v rovnakom segmente sčítali. Ďalej budeme pracovať práve s údajmi o celkovej hmotnosti predaných výrobkov, kvôli vyššie zmieneným nevýhodám prvých dvoch možností nazerania na dáta. Aby sme nepracovali s príliš veľkými číslami, tak celkovú hmotnosť budeme uvažovať v tonách.

Každý segment obsahuje sušienky zaradené do užších podskupín, tzv. brandov. Brandy sú v jednotlivých segmentoch zastúpené v rôznych podieloch. V prípade, že jeden brand má v segmente značnú proporcionálnu prevahu, tak sa sezónnosť bude riadiť jeho predajom a jeho promočnými aktivitami. Z tohto dôvodu upravíme proporcionalitu brandov v každom segmente tak, že zvolíme normovanie na 100 ton za celú dobu štúdie.

Pri zbieraní dát boli celé týždne zaradené do toho mesiaca, do ktorého pripadal pondelok v danom týždni. Preto upravíme ešte dĺžky mesiacov tak, aby mali rovnaký počet dní. Budeme uvažovať, že každý mesiac má 30.43 dní, čo je priemerný počet dní v mesiaci v priebehu rokov zbierania dát. Urobíme korekciu mesačných údajov tak, že celkovú hmotnosť za konkrétny mesiac vydělíme jeho aktuálnym počtom dní a vynásobíme nami uvažovaným počtom. Tým predídeme možným väčším výkyvom v údajoch v jednotlivých mesiacoch.

Ďalej budeme analyzovať celkovo jedenásť časových radov, pričom jeden časový rad odpovedá jednému segmentu sušienok, ktoré označíme ako segment A až segment K. Dĺžka každého časového radu bude úmerná počtu meraní, pričom máme jedno meranie za mesiac od augusta 2004 do októbra 2008, t.j. 51 meraní.

Pre lepšie porozumenie chovania časového radu si najprv urobíme jeho rozklad na jednotlivé zložky. Rozklad môže byť aditívny alebo multiplikatívny. V prípade aditívneho rozkladu sú jednotlivé zložky časového radu v rovnakých merných jednotkách ako pôvodný rad a používa sa, keď variabilita hodnôt časového radu je približne konštantná v čase. Po multiplikatívnom rozklade je iba trendová zložka časového radu v rovnakých merných jednotkách ako pôvodný časový rad a ostatné zložky sú v relatívnom vyjadrení voči nej. Tento spôsob rozkladu sa používa v prípade, že variabilita časového radu rastie



v čase, alebo sa v čase mení, ako je to v prípade našich časových radov. Budeme teda predpokladať, že každý náš časový rad má dekompozíciu tvaru

$$y_t = Tr_t Sz_t E_t,$$

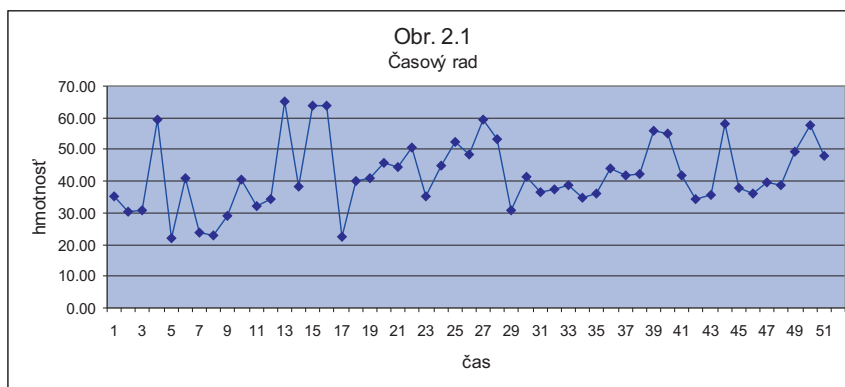
kde  $Tr_t$  označuje trendovú zložku,  $Sz_t$  sezónnu zložku,  $E_t$  náhodnú zložku a  $t$  čas. Odhady zložiek budeme ďalej značiť malými začiatocnými písmenami.

Náhodnú (reziduálnu) zložku si môžeme predstaviť ako postupnosť náhodných veličín tvorenú náhodnými pohybmi v časovom rade. Jej charakter sa nedá rozpoznať a pokrýva chyby pri meraní údajov. Budeme predpokladať, že náhodná zložka je biely šum, t.j. proces nekorelovaných náhodných veličín s nulovou strednou hodnotou a rovnakým konečným rozptylom.

V ďalších podkapitolách sa budeme podrobnejšie zaoberať trendom a sezónnou zložkou.

## 2.2 Trend

Trend vyjadruje dlhodobé zmeny v priemernom chovaní časového radu. Vzniká pôsobením rôznych faktorov ako sú zmeny vo výške priemernej mzdy, rast trhu, zmeny požiadaviek kupujúcich a iné. Vďaka trendu je možné predikovať budúce správanie časového radu alebo presnejšie konštruovať jeho sezónnu zložku.



Na obrázku 2.1 je zobrazený časový rad segmentu sušienok A. Z grafu je zrejmé, že rad postupne v čase mení globálne svoj charakter. To znamená, že na popis trendu nemôžeme použiť metódu preloženia radu nejakou konkrétnou krivkou ale metódu, ktorá

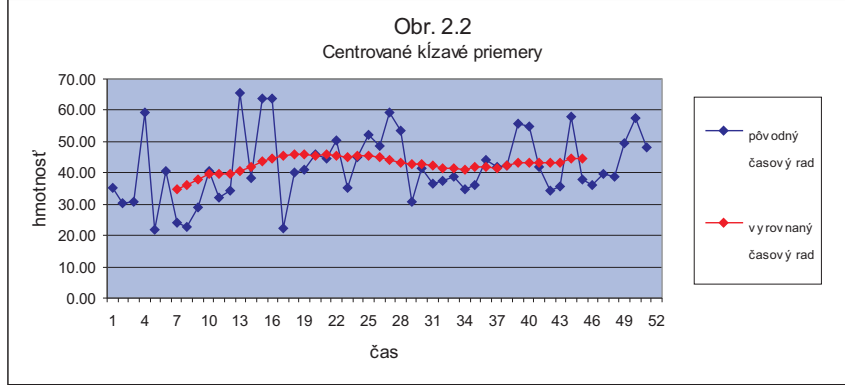
bude trend vyrovňovať postupne na intervaloch. Pri každom časovom rade použijeme metódu centrovaných klzavých priemerov, ktorá je výpočetne jednoduchá, dostatočne presná a v praxi často používaná.

Klzávé priemery sú lineárne kombinácie členov pôvodného radu. Pri jednoduchých klzavých priemeroch toto vyjadrenie odpovedá vyrovnaniu krátkych úsekov radu lineárnym alebo konštantným trendom. Najprv musíme vhodne zvoliť dĺžku priemerov, pričom tá by mala byť rovnako dlhá ako perióda sezónnych výkyvov. Pri voľbe dĺžky 12, čo odpovedá dĺžke periódy, by ale priemer dvanástich mesiacov spadol medzi dva ďalšie mesiace a nevedeli by sme, ku ktorému z nich ho priradiť. Preto volíme dĺžku 13 a dostaneme centrované klzávé priemery tvaru

$$\begin{aligned} tr_t &= \frac{1}{2} \left( \frac{1}{12} (y_{t-6} + y_{t-5} + \dots + y_{t+5}) + \frac{1}{12} (y_{t-5} + y_{t-4} + \dots + y_{t+6}) \right) \\ &= \frac{1}{24} (y_{t-6} + 2y_{t-5} + 2y_{t-4} + \dots + 2y_{t+4} + 2y_{t+5} + y_{t+6}). \end{aligned}$$

Tabuľka 2.1																	
rok	2004			2005			2006			2007			2008				
mes.	t	y <sub>t</sub>	tr <sub>t</sub>	t	y <sub>t</sub>	tr <sub>t</sub>	t	y <sub>t</sub>	tr <sub>t</sub>	t	y <sub>t</sub>	tr <sub>t</sub>	t	y <sub>t</sub>	tr <sub>t</sub>		
Jan	1	35.17		6	40.75		18	40.06	45.86	30	41.54	42.84	42	34.12	43.34		
Feb				7	23.98	34.69	19	40.83	45.76	31	36.34	42.37	43	35.84	43.44		
Mar				8	22.76	36.28	20	46.01	45.65	32	37.39	41.68	44	57.95	44.38		
Apr				9	28.98	37.98	21	44.53	45.84	33	38.80	41.29	45	37.78	44.68		
Máj				10	40.63	39.54	22	50.46	45.27	34	34.62	41.20	46	36.30			
Jún				11	32.12	39.75	23	35.17	45.20	35	36.11	41.71	47	39.53			
Júl				12	34.31	39.74	24	44.96	45.62	36	44.08	41.86	48	38.86			
Aug				13	65.32	40.41	25	52.31	45.50	37	42.00	41.53	49	49.44			
Sep				2	30.36		14	38.30	42.08	26	48.60	44.95	38	42.48	42.36	50	57.63
Okt				3	30.97		15	63.81	43.70	27	59.34	44.35	39	55.93	43.10	51	48.03
Nov				4	59.29		16	63.96	44.75	28	53.48	43.45	40	54.83	43.21		
Dec				5	21.93		17	22.27	45.29	29	30.97	42.83	41	41.89	43.42		

V tabuľke 2.1 sú uvedené hodnoty časového radu segmentu sušienok A a ich klzávé priemery a na obrázku 2.2 je zobrazený pôvodný a vyrovnaný časový rad.



Z tabuľky aj grafu vyššie je vidno, že nami použité klzavé priemery nevyrovnávajú prvých a posledných 6 hodnôt časového radu, čo je dané metodikou jeho vyrovnávania. Táto skutočnosť je pre nás ale nezaujímavá, pretože nemá významný vplyv na ďalšiu analýzu a konštrukciu sezónnej zložky.

Po aplikácii klzavých priemerov získavame očistené časové rady od sezónnej zložky. Z grafov týchto radov môžeme už lepšie rozpoznať ich charakter v priebehu času, či už sa jedná o jeho dlhodobý pokles alebo nárast. Pre lepšie znázornenie je však ešte možné preložiť vyrovnaný rad nejakou konkrétnou krivkou.

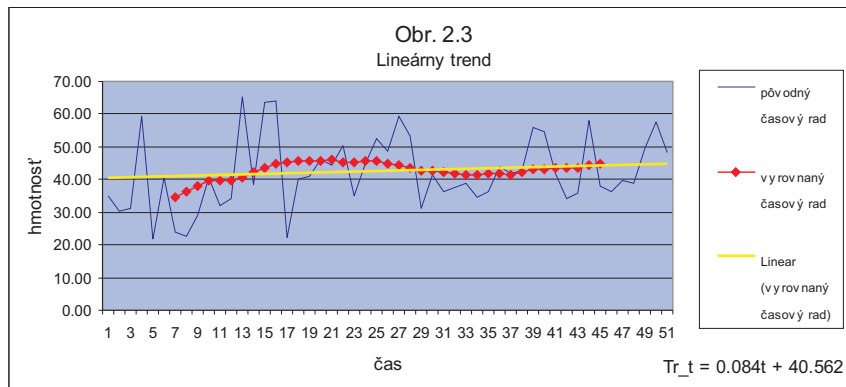
Z pohľadu na vyrovnaný rad na obrázku 2.2 vidíme, že ho môžeme preložiť priamkou. Jedná sa teda o lineárny trend v tvare

$$Tr_t = \beta_0 + \beta_1 t,$$

kde  $\beta_0$  a  $\beta_1$  sú parametre, ktoré odhadneme pomocou sústavy rovníc

$$\begin{aligned} \hat{\beta}_0 + \hat{\beta}_1 \sum_{t=1}^n t &= \sum_{t=1}^n y_t, \\ \hat{\beta}_0 \sum_{t=1}^n t + \hat{\beta}_1 \sum_{t=1}^n t^2 &= \sum_{t=1}^n t y_t, \end{aligned}$$

pričom  $n$  značí počet pozorovaní a  $\hat{\beta}_0$  a  $\hat{\beta}_1$  odhady  $\beta_0$  a  $\beta_1$ .



Na obrázku 2.3 je zobrazený časový rad so zodpovedajúcim lineárnym trendom.

## 2.3 Sezónnosť

Sezónna zložka popisuje pravidelne sa opakujúce zmeny v časovom rade, ktoré sa odohrávajú počas jedného roka. Táto zložka vzniká ako dôsledok niektorých faktorov, napríklad striedanie ročných období. Na trhu sušienok a oblátiek hrá sezónnosť dôležitú úlohu, a preto ju budeme analyzovať.

Na eliminovanie sezónnej zložky zvolíme jednoduchú ale účinnú metódu, pri ktorej použijeme už vypočítané centrované kľzavé priemery. Budeme pritom predpokladať, že naše časové rady majú pravidelný časový priebeh, to znamená, že sezónne faktory sú každý rok rovnaké. Pomocou centrovaných kľzavých priemerov najprv vypočítame odhad sezónnej zložky neočistenej od náhodných výkyvov ako

$$sz_t e_t = y_t / tr_t.$$

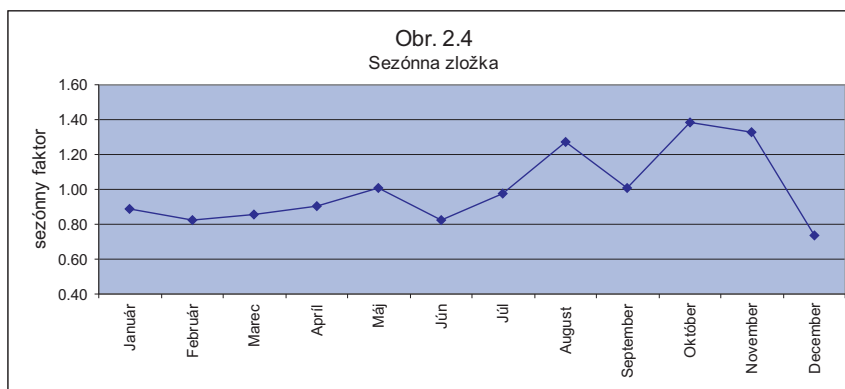
Náhodnú zložku odstránime tak, že spriemerujeme odhady za rovnaké mesiace. Takýto priemer  $s\tilde{z}_t$  za každý mesiac ďalej znormalizujeme pomocou vzťahu

$$sz_t = \frac{12s\tilde{z}_t}{\sum_{t=1}^{12} s\tilde{z}_t}$$

a tým dostaneme už požadované odhady sezónnych faktorov.

Tabuľka 2.2						
mesiac	$sz_t e_t$				priemery $sz_t e_t$	$sz_t$
Jan	0.87	0.97	0.79		0.88	0.88
Feb	0.69	0.89	0.86	0.83	0.81	0.82
Mar	0.63	1.01	0.90	1.31	0.84	0.85
Apr	0.76	0.97	0.94	0.85	0.89	0.90
Máj	1.03	1.11	0.84		0.99	1.01
Jún	0.81	0.78	0.87		0.82	0.83
Júl	0.86	0.99	1.05		0.97	0.98
Aug	1.62	1.15	1.01		1.26	1.27
Sep	0.91	1.08	1.00		1.00	1.01
Okt	1.46	1.34	1.30		1.36	1.38
Nov	1.43	1.23	1.27		1.31	1.33
Dec	0.49	0.72	0.96		0.73	0.74

V tabuľke 2.2 je uvedená konkrétna konštrukcia sezónnej zložky pre skupinu sušienok A a na obrázku 2.4 jej graf.



**Poznámka ku predikcii.** Pri analýze časových radov sa často konštruuje predpoveď budúcich hodnôt. Pre trh rýchloobrátkového tovaru je však prevádzanie predikcie nepodstatné, pretože naň má vplyv priveľa faktorov, ako sú rôzne akcie, vstup nových výrobkov na trh a sťahovanie starých. Preto sa predikciou nebudeme bližšie zaoberať, iba zmienime jeden z postupov. Krivku prekladajúcu očistený rad od sezónnej zložky, viď obrázok 2.3, by sme predĺžili na požadovaný interval a v jednotlivých predpovedných okamihoch by sme vynásobili príslušnú hodnotu odpovedajúcim sezónnym faktorom.

## 2.4 Interpretácia výsledkov

Cieľom analýzy časových radov je lepšie porozumenie mechanizmu správania sledovaného javu. V našom prípade sa jednalo o analýzu trhu rýchloobrátkového tovaru sušienok a oblátiek, vďaka ktorej sme schopní celkom presne popísať jeho chovanie v čase, či už sa jedná o dlhodobý trend alebo sezónne správanie počas roka.

Pri interpretácii výsledkov našej analýzy by sme mali zobrať do úvahy niektoré faktory, ktoré mohli prispieť k skresleniu týchto výsledkov. Je zrejmé, že v priebehu vyše štyroch rokov, počas ktorých sa dáta zoskupovali, dochádzalo k ekonomickým zmenám na trhoch. Avšak ekonomická situácia od augusta 2004 až do októbra 2008 bola v Českej republike relatívne stabilná, a preto významne nevplývala na nami skúmaný trh a bližšie sme ju nezohľadňovali. Naopak, informácie o priemernej cene a hmotnosti za celé obdobie štúdie, ktoré sme mali k dispozícii, nám síce zjednodušili prácu s dátami, ale na druhej strane sme neboli schopní určiť promočné aktivity výrobu (akciové ceny a pod.) v konkrétnom časovom okamihu merania. To mohlo mať za príčinu náhly nárast predaja jedných druhov výrobkov a pokles predaja iných druhov. Výkyvy v predaji sušienok rôznych segmentov mohli byť takisto spôsobené skutočnosťou, že niektoré výrobky vstupovali na trh v rôznych mesiacoch štúdie, prípadne sa na ňom po celé toto obdobie neudržali.

Vplyv týchto javov sme čiastočne potlačili tým, že sme výrobky skumulovali do väčších skupín, segmentov. V každom segmente sme mali ešte rozdelenie sušienok do brandov, z ktorých mali niektoré značnú prevahu v celkovej predanej hmotnosti. Preto, aby sa sezónnosť riadila predajom všetkých koncových výrobkov, tak sme upravili vnútri každého segmentu proporionalitu brandov.

Pre každý segment sušienok A až K sme skonštruovali trendovú a sezónnu zložku. Ich grafy sú vykreslené v Prílohe 1, postupne pre každý segment.

Prvý graf pri každej skupine sušienok zobrazuje jej časový rad a príslušný trend. Až na jednu výnimku je z grafov zrejmé, že trend dlhodobo rastie. Ďalej môžeme pozorovať, že trend v prípade segmentov A, E, F alebo I rastie pomaly v čase. Naopak, trendové zložky segmentov C, D, G, H, J a K majú výraznejšiu rastúcu tendenciu. V prípade sušienok zo segmentu B je ale situácia opačná. Pre lepšie znázornenie sme cez vyrovnaný rad ešte preložili priamku, čím sme získali lineárny trend. Z toho je už jasne vidno, že ide o pokles narozdiel od ostatných segmentov.

Druhý graf príslušiaci k jednotlivým segmentom vykresluje sezónnu zložku. Vďaka nej vieme povedať, v ktorých mesiacoch v roku sa daná skupina sušienok najlepšie pre-

dáva a v ktorých sa naopak predáva najhoršie. Sezónne faktory v rôznych segmentoch môžu byť na sebe závislé, pretože napríklad v letných mesiacoch si spotrebitelia kupujú sušienky z jedného segmentu na úkor sušienok z druhého a v zime je to opäť naopak. Na sezónnej zložke segmentu A vidíme, že jej predaj je od januára do júla v podstate nemenný, pričom od augusta do novembra sa predaj výrazne zvýši. Segmenty B a J majú najlepší odbyt v máji a októbri, segmenty G a I v septembri. Skupina sušienok C postupne od januára predaj zvyšuje až do novembra, skupine E predaj rastie do júla a odtiaľ zas postupne klesá. V segmentoch D, H a K je predaj počas roka porovnateľný, až na posledný mesiac december, kedy sa predaj výrazne prepadne. Nízky dopyt v decembri si môžeme vysvetliť napríklad tým, že v okolí Vianoc ľudia pečú vlastné koláče a sušienky. Tento decembrový prepád predaja sledujeme u všetkých skupín okrem skupiny F, u ktorej je naopak december najúspešnejší mesiac čo sa týka predajnosti.

Získané informácie o trende a sezónnosti jednotlivých segmentov sušienok sú veľmi užitočné. Výrobcom a predajcom konkrétnych produktov môžu pomôcť pri plánovaní ďalšej výroby a určovaní efektívnejšej stratégie predaja.

# Kapitola 3

## Priestorová závislosť

Pri analýze dát z oblasti trhu rýchloobrátkového tovaru nás obvykle popri časovej závislosti zaujíma aj závislosť priestorová. Výška odbytu výrobkov v konkrétnych oblastiach krajiny môže byť závislá na rôznych faktoroch, ako sú priemerný vek obyvateľstva, miera nezamestnanosti, priemerná mzda a s tým súvisiaca životná úroveň v danom kraji a iné.

V tejto kapitole budeme študovať predajnosť sušienok rôznych cenových hladín v jednotlivých krajoch Českej republiky celkovo za celé obdobie štúdie. Bude nás zaujímať závislosť medzi životnou úrovňou v konkrétnom kraji a cenou, ktorú sú ochotní kupujúci za výrobok v tomto kraji zaplatiť. Takisto budeme overovať, či spotrebitelia uprednostňujú cenu pred kvalitou, t.j. či kupujú vždy najlacnejšie výrobky. Všetky analýzy budeme prevádzať na jednom, vhodne vybranom segmente sušienok (segmente D), to znamená na segmente, ktorý obsahuje dostatočne veľký počet sušienok rôznych cenových hladín.

### 3.1 Úprava dát

Najprv, ešte predtým ako začneme s analýzou dát, sa musíme rozhodnúť a prehodnotiť, pre ktorý segment ju budeme robiť. V segmentoch budeme uvažovať len tie výrobky, ktoré sa predávali vo všetkých štrnástich krajoch v rovnakom období. Túto úpravu dát je potrebné urobiť, aby porovnanie predajnosti medzi kraji malo dobrú výpovednú hodnotu a nebude predstavovať dramatický zásah do dát, keďže veľké obchodné reťazce majú podobný sortiment. V každom segmente výrobky najprv prevedieme na rovnakú gramáž 100 gramov tak, že ich cenu vydelíme ich pôvodnou hmotnosťou a potom



vynásobíme hodnotou požadovanej gramáže 100. Toto je opäť len jedna z možností nazerania na dáta. Mohli by sme skúmať napríklad cenu za kus a nebrať ohľad na odlišnú hmotnosť výrobkov. Aj keď je pravda, že pri tovare ako sú sušienky sú rozdiely medzi hmotnosťami väčšinou minimálne a spotrebiteľ sa preto podľa nich neorientuje, tak sme zvolili úpravu na rovnakú gramáž pre všetky výrobky. Pretože každý segment obsahuje výrobky rovnakého alebo podobného typu, čiže substitúty, tak ich prevedením na rovnakú gramáž získame informáciu o skutočných rozdieloch medzi hladinami cien daných sušienok. Segment, s ktorým chceme pracovať, by mal obsahovať sušienky rôznych cenových hladín a čo najväčší rozdiel v cenách. To znamená, že pre každý segment vypočítame rozptyl cien a nakoniec budeme analyzovať ten s najväčším rozptylom. Konkrétne budeme počítat výberový rozptyl  $S_n^2$  a to pomocou vzťahu

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

kde  $n$  značí počet uvažovaných sušienok v segmente,  $X_i$  cenu  $i$ -tej sušienky v segmente a  $\bar{X}_n$  výberový priemer cien v segmente vypočítaný ako

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Vo vybranom segmente ďalej zoskupíme sušienky do troch skupín podľa ich ceny. Budeme teda porovnávať predaj sušienok s najnižšími (budeme značiť L), najvyššími (D) a priemernými cenami (P).

Po aplikovaní vyššie popísaného postupu pre výrobky vo vybranom segmente spočítame celkový počet kusov predaných za celé obdobie zberu dát, zvlášť pre všetky kraje. Pre každý výrobok túto sumu vynásobíme jeho skutočnou hmotnosťou. Tým získame údaj o celkovej predanej hmotnosti výrobku, ktorý budeme ďalej analyzovať, podobne ako v druhej kapitole. Vo vytvorených troch cenových podskupinách sčítame celkové hmotnosti daných sušienok a tým získame pre každý kraj údaje o množstve predaných výrobkov týchto troch skupín. Avšak každá cenová skupina môže obsahovať iný počet výrobkov, a preto upravíme ich proporcionalitu, aby sme porovnávali rovnocenné skupiny.

Kraje Českej republiky sú rôzne veľké a takisto majú rôzny počet obyvateľov, čo sa odráža v konečnom zakúpenom množstve pre konkrétne kraje. Preto poslednou úpravou, ktorú pred analýzou prevedieme je prepočítanie množstva predaných sušienok v závislosti na počte obyvateľov v jednotlivých krajoch. K dispozícii máme údaje

Českého statistického úřadu o počte obyvatelů v krajích za roky studie 2004 až 2008. Tieto hodnoty spriemerujeme a údaje o predaných množstvách vydelíme v každom kraji výsledným priemerom. Pretože dáta, s ktorými pracujeme, sú zoskupením údajov len z medzinárodných obchodných reťazcov a nie zo všetkých obchodov Českej republiky, tak touto úpravou nezískame presnú hodnotu zakúpeného množstva na jedného obyvateľa. Pre jednoduchosť však budeme ďalej tento údaj označovať ako hmotnosť predaných sušienok na jedného obyvateľa a zmienený fakt budeme brať do úvahy takisto ako fakt, že obchodné reťazce sú v krajoch zastúpené v rozdielnych pomeroch.

Po úprave dát máme nakoniec k dispozícii údaje o celkovej hmotnosti predaných výrobkov v gramoch na jedného obyvateľa jednotlivých krajov, a to ako pre jednotlivé výrobky, tak aj pre vytvorené tri cenové hladiny sušienok. Tieto informácie budeme v ďalších podkapitolách študovať a tiež budeme analyzovať ich závislosť na mieste predaja, t.j. konkrétnych krajoch Českej republiky.

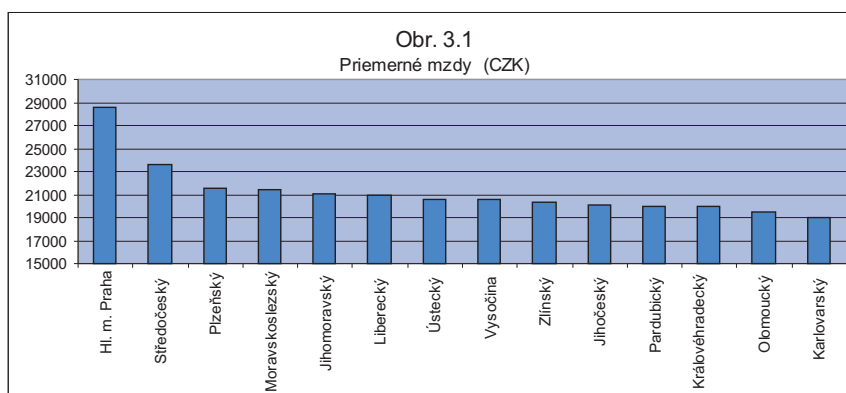
Pre jednoduchosť budeme ďalej predpokladať, že medzi kraji neexistuje žiadna vzájomná závislosť (napr. prechádzaním do susedných krajov sa postupne mení životná úroveň), budeme si ich predstavovať usporiadané v rade bez akejkoľvek náväznosti.

## 3.2 Popisná štatistika

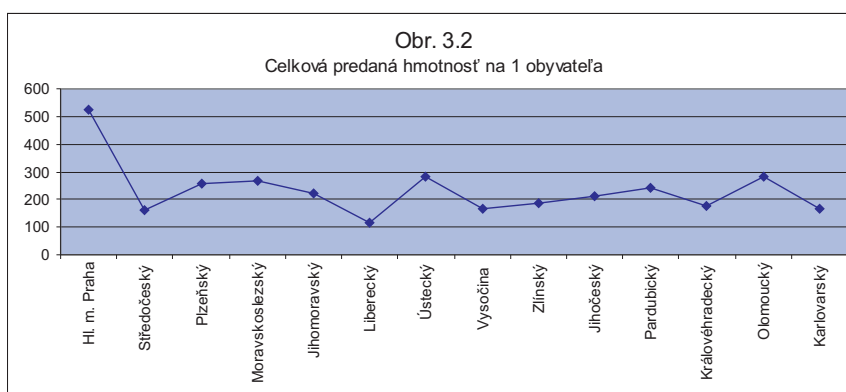
Popisná štatistika sa používa na analýzu hromadných dát, čím získame ich prehľadnejší a úspornejší popis a tým aj popis sledovaného výseku reálneho sveta, v našom prípade trhu rýchloobrátkového tovaru. Vzhľadom k tomu, že získané dáta sú ale v surovej podobe len neusporiadanou a chaotickou hromadou údajov, tak z nich bez ďalšieho spracovania nevieme vyčítať prakticky žiadne užitočné informácie. Spôsob spracovania a úpravu dát sme opísali v prvej časti tretej kapitoly, a preto sa v tejto časti zameriame hlavne už na popis konkrétnych výsledkov vyplývajúcich z našich dát, teda na popisnú štatistiku.

Segment sušienok, s ktorým nakoniec pracujeme a ktorý sme podľa už skôr zmieneného postupu vybrali je segment D. Táto skupina obsahuje dostatočne veľa výrobkov, ktoré sa predávali vo všetkých krajoch po celú dobu štúdie. Výrobky majú takisto dostatočne veľký cenový rozptyl, čo môžeme vidieť aj v prvej tabuľke Prílohy 2, ktorá zobrazuje tieto výrobky, ich ceny a príslušnosť do konkrétnej cenovej skupiny. Výrobok pritom spadá do cenovej skupiny L, ak jeho cena za 100 gramov je nižšia ako desať českých korún, do skupiny P, ak sa jeho cena pohybuje medzi desiatimi až dvadsiatimi korunami a nakoniec výrobky nad dvadsať korún spadajú do cenovej skupiny D.

Jeden z hlavných faktorov určujúcich životnú úroveň v kraji je výška priemernej mzdy. Opäť sme mali k dispozícii údaje Českého statistického úradu o priemernej mzde vo všetkých krajoch Českej republiky za roky 2004 až 2008. Pretože nás ale nezaujímá predaj za jednotlivé roky, teda časová závislosť ale priestorová, tak sme z týchto údajov urobili aritmetický priemer, čím sme dostali priemernú mzdu za celé štyri roky od roku 2004 do 2008.



Na obrázku 3.1 je graf zobrazujúci výšku priemernej mzdy, postupne od kraja s najvyššou mzdou po kraj s najnižšou. Nebudeme uvádzať konkrétne tabuľkové hodnoty, pretože pre nás nie sú až také dôležité, ale dôležitý je pomer výšok miezd medzi jednotlivými kraji. Pre lepšie porovnanie životnej úrovne a množstva predaných sušienok sú v ďalších grafoch kraje usporiadané ako na obrázku 3.1, to znamená od kraja s najvyššou priemernou mzdou, t.j. hlavné mesto Praha, po ten s najnižšou, Karlovarský kraj. Obrázok 3.2 zobrazuje hmotnosť predaných sušienok všetkých cenových hladín v segmente D.



Z pohľadu na obrázok 3.2 môžeme usúdiť, že trend dopytu po sušienkach s poklesom životnej úrovne mierne klesá. Vidíme, že najväčšie množstvo na jedného obyvateľa sa predalo v Prahe a to so značným nárastom a najmenej v Libereckom kraji.

V ostatných krajoch tieto rozdiely nie sú veľmi výrazné. Skutočnosť, že v niektorých krajoch sa predá viac sušienok ako v iných môže byť tiež spôsobená rôznym zastúpením obchodných reťazcov a odlišným životným štýlom v jednotlivých krajoch, prípadne väčším dopytom v ostatných segmentoch. Takisto aj najväčší odbyt v hlavnom meste Praha je pravdepodobne spôsobený tým, že sú v ňom zastúpené všetky obchodné reťazce, ktoré nám dodali dáta, a v najväčšom pomere.

V prípadoch skúmania predaja sušienok troch rozličných cenových skupín sa grafy správajú podobne ako na obrázku 3.2. Príslušné grafy sa nachádzajú v Prílohe 2 a je z nich zrejmé, že najviac predávaných sušienok je z cenovej skupiny L. Grafy cenových skupín sme ďalej preložili konštantnou funkciou, ktorá vyjadruje priemer predanej hmotnosti v danej cenovej skupine zo všetkých krajov okrem Prahy.

Tabuľka 3.1						
Kraj	L	Pomer	D	Pomer	P	Pomer
Stredočeský	98.38	0.830	8.11	0.699	55.06	0.685
Plzeňský	146.90	1.240	15.17	1.307	93.37	1.161
Moravskoslezský	158.72	1.340	14.89	1.283	91.88	1.143
Jihomoravský	119.02	1.004	14.76	1.271	89.88	1.118
Liberecký	54.98	0.464	7.05	0.607	54.60	0.679
Ústecký	175.25	1.479	13.80	1.189	92.69	1.153
Vysočina	90.87	0.767	7.85	0.676	66.94	0.833
Zlínský	73.30	0.619	7.61	0.655	104.69	1.302
Jihočeský	120.64	1.018	11.07	0.953	79.23	0.986
Pardubický	134.20	1.133	12.68	1.092	93.13	1.158
Královéhradecký	107.42	0.907	9.79	0.843	61.07	0.760
Olomoucký	159.03	1.342	17.19	1.481	108.20	1.346
Karlovarský	101.60	0.857	10.97	0.945	54.36	0.676
Priemerná hmotnosť	118.49		11.61		80.39	

V tabuľke 3.1 sú konkrétne hodnoty hmotností predaných sušienok a spomínaný priemer predanej hmotnosti pre každú cenovú skupinu. V tabuľke je ďalej pre každý kraj vypočítaný pomer medzi skutočnou a priemernou predanou hmotnosťou. Aby bol priemer a následne aj pomer týchto údajov viac vypovedajúci, tak sme do neho nezapočítavali hodnoty zozbierané v Prahe, čo je vidno z grafov aj príslušnej tabuľky. Dôvodom je práve veľký rozdiel medzi Prahou a ostatnými krajmi v množstve predaných sušienok, ktorý môžeme pozorovať na obrázku 3.2, a ktorý je spôsobený práve väčším zastúpením reťazcov. Porovnaním hodnôt v tabuľke zisťujeme, že najväčší pomer cenovej skupiny L lacných výrobkov a jej priemeru dosahuje Ústecký kraj a v cenovej skupine luxusnejšieho tovaru D aj v skupine P je to Olomoucký kraj. Prekvapivé je zistenie, že v Olomouckom kraji, s druhou najnižšou priemernou mzdou,

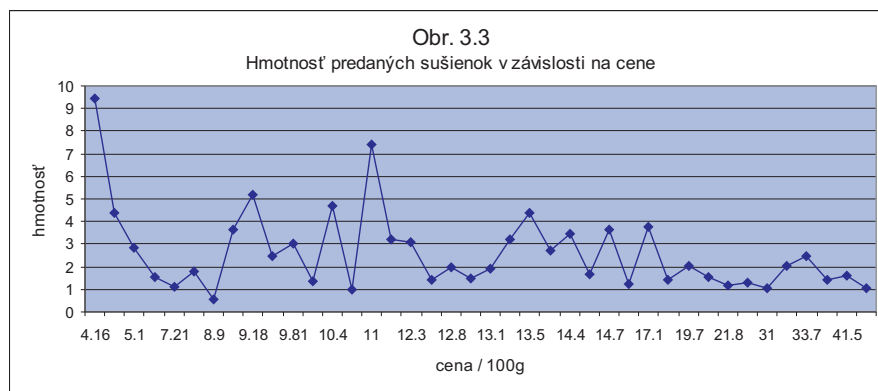
kupujú spotrebitelia sušienky cenovej skupiny D nad priemer celej Českej republiky a ešte v najväčšom pomere. Naopak Stredočeský kraj, s druhou najvyššou priemernou mzdou, nedosahuje ani priemer pre Českú republiku a to v žiadnej z cenových skupín. To je ale možný následok toho, že obyvatelia Stredočeského kraja jazdia nakupovať do Prahy. Najnižší pomer skutočnej a priemernej predanej hmotnosti vo všetkých troch cenových skupinách má Liberecký kraj. Kraje Vysočina, Stredočeský, Královéhradecký a Karlovarský sú vo všetkých cenových skupinách pod priemerom, zatiaľ čo Plzeňský, Moravskoslezský, Jihomoravský, Ústecký, Pardubický a Olomoucký kraj dosahujú opäť v každej skupine vyššie hodnoty ako je priemer danej cenovej skupiny. Zlínsky kraj je nad priemerom iba v skupine sušienok s priemernou cenou a Jihočeský v skupine lacných výrobkov.

Tabuľka 3.2			
Kraj	L	D	Pomer L/D
Hl. m. Praha	315.81	33.04	9.558
Stredočeský	98.38	8.11	12.131
Plzeňský	146.90	15.17	9.684
Moravskoslezský	158.72	14.89	10.660
Jihomoravský	119.02	14.76	8.064
Liberecký	54.98	7.05	7.799
Ústecký	175.25	13.80	12.699
Vysočina	90.87	7.85	11.576
Zlínský	73.30	7.61	9.632
Jihočeský	120.64	11.07	10.898
Pardubický	134.20	12.68	10.584
Královéhradecký	107.42	9.79	10.972
Olomoucký	159.03	17.19	9.251
Karlovarský	101.60	10.97	9.262

V tabuľke 3.2 máme opäť po jednotlivých krajoch predané hmotnosti sušienok z cenových hladín L a D a v poslednom stĺpci je ich pomer. Pozeráme sa teda na rozdiel pomeru cenových skupín v jednotlivých krajoch, konkrétne na pomer predaných sušienok z cenovej skupiny L lacných výrobkov k luxusnejšiemu tovaru skupiny D. Porovnaním hodnôt v tabuľkách 3.1 a 3.2 získame lepšiu informáciu o tom, v ktorom kraji sa riadia spotrebitelia viac cenou a v ktorom kvalitou. Ako už vieme, najviac sušienok na osobu sa predá v Prahe a to platí pre všetky tri cenové skupiny. Navyše má Praha najvyššiu priemernú mzdu zo všetkých krajov Českej republiky, a preto sme očakávali, že kupujúci budú uprednostňovať kvalitu najviac zo všetkých krajov a teda pomer najlacnejších výrobkov k najdrahším bude práve tu najnižší. Tabuľka 3.2 ale toto tvrdenie vyvracia. Najmenší pomer je v Libereckom kraji, kde spotrebitelia kupujú lacnejšie sušienky o necelý osemnásobok hmotnosti drahších. Potom nasleduje Jihomoravský kraj, za ním Olomoucký a Karlovarský s najnižšou priemernou mzdou a až potom Praha. Najvyššie hodnoty dosahuje Ústecký a hneď po ňom Stredočeský

kraj, kde predaná hmotnosť lacných výrobkov je dvanásťkrát väčšia ako tých z cenovej skupiny D. Po týchto zisteniach môžeme teda tvrdiť, že výška priemernej mzdy nemá na trh medzinárodných reťazcov so sušienkami alebo aspoň na študovaný segment D výrazný vplyv, ale skôr naň pôsobia iné faktory.

Posledné na čo sa zameriame je predajnosť všetkých výrobkov od najlacnejšieho po najdrahší. Pre všetky kraje sme vykreslili grafy zobrazujúce hmotnosť predaných sušienok na jedného obyvateľa v závislosti na ich cene (viď Príloha 2) a na obrázku 3.3 už vidíme konkrétny graf pre Prahu.



Z grafu je zrejmé, že so stúpajúcou cenou výrobkov, až na výnimky, klesá ich zakúpená hmotnosť. Toto platí pre všetky grafy, a teda aj pre všetky kraje. Výnimku tvorí hneď najlacnejší výrobok, ktorého sa v Prahe predalo najviac zo všetkých, ale v ostatných krajoch má malé zastúpenie. Ostatné prípady, kedy zakúpená hmotnosť drahšieho výrobku je vyššia ako lacnejšieho, sú vo všetkých krajoch zhodné, a preto predpokladáme, že sa jedná medzi kupujúcimi o osvedčené a obľúbené produkty. Otázkou atraktivity lacnejších výrobkov sa budeme ďalej zaoberať v poslednej časti kapitoly.

### 3.3 Lineárna regresia

Cieľom regresnej analýzy je vyšetrovanie vzťahu dvoch alebo viacerých premenných, spoznať pravdepodobnú formu tohto vzťahu a odhadnúť, resp. predpovedať hodnotu odpovedajúcu inej hodnote. V prípade dvoch premenných, hovoríme o jednoduchom lineárnom regresnom modeli a jeho aplikáciou na naše dáta overíme hypotézy, či kupujúci uprednostňujú lacnejšie výrobky oproti tradičným drahším. Najprv veľmi stručne

popíšeme model jednoduchkej lineárnej regresie a následne ukážeme jeho použitie.

Nech  $n \geq 3$  je počet pozorovaní. Majme vektor náhodných veličín  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ , nazývame ich závislé premenné, a maticu daných čísel  $\mathbf{X} = (X_{ij})$  typu  $n \times 2$ , ktoré nazývame nezávislé premenné. V prípade jednoduchkej lineárnej regresie platí  $X_{i1} = 1$  pre všetky  $i$  a za splnenia určitých predpokladov má tvar

$$Y_i = \beta_0 + \beta_1 X_{i2} + e_i, \quad i = 1, \dots, n,$$

kde  $\beta_0$  a  $\beta_1$  sú neznáme parametre a  $e_1, \dots, e_n$  sú nezávislé náhodné veličiny s normálnym rozdelením pravdepodobností  $N(0, \sigma^2)$ . Pre zjednodušenie budeme namiesto  $X_{i2}$  ďalej písať len  $X_i$ .

Predpoklady, na ktorých je ďalej model lineárnej regresie založený, sú:

1. vektor  $(X_1, \dots, X_n)^T$  neobsahuje rovnaké prvky, a teda stĺpce matice  $\mathbf{X}$  sú lineárne nezávislé,
2. náhodná veličina má strednú hodnotu  $EY_i = \beta_0 + \beta_1 X_i$  a konštantný rozptyl  $\text{var}Y_i = \sigma^2 > 0$ ,
3.  $Y_1, \dots, Y_n$  sú nezávislé.

Na prevedenie testov štatistických hypotéz s regresnými koeficientmi potrebujeme poznať najprv odhady regresných koeficientov a rozptylu. Parametre  $\beta_0$  a  $\beta_1$  sa odhadujú metódou najmenších štvorcov, t.j. minimalizáciou výrazu

$$(\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta),$$

kde  $\beta = (\beta_0, \beta_1)^T$ . Zložitejším odvodzovaním, ktoré ale nebudeme uvádzať, sa dá ukázať, že odhady našich parametrov sú dané vzťahmi

$$\hat{\beta}_0 = \frac{\sum_{i=1}^n X_i^2 \sum_{i=1}^n Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n X_i Y_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2},$$

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2}.$$

Taktisto nebudeme odvodzovať odhad rozptylu  $\sigma^2$ , ale rovno uvedieme vzťah pre jeho výpočet a ten je nasledovný

$$s^2 = (\sum_{i=1}^n Y_i^2 - \hat{\beta}_0 \sum_{i=1}^n Y_i - \hat{\beta}_1 \sum_{i=1}^n X_i Y_i) / (n - 2).$$

Parameter  $\beta_0$  vyjadruje posun regresnej priamky oproti nule. Preto sa zaoberáme častejšie parametrom  $\beta_1$ , ktorý predstavuje prírastok pri zmene nezávislej premennej  $X_i$ . Pre test hypotézy  $H_0 : \beta_1 = 0$  proti alternatíve  $H_1 : \beta_1 \neq 0$  na hladine  $\alpha$  je kritický obor

$$|\hat{\beta}_1| \sqrt{\sum_{i=1}^n X_i^2 - n\bar{X}^2}/s \geq t_{n-2}(\alpha),$$

kde  $\bar{X} = \sum_{i=1}^n X_i/n$  a  $t_{n-2}(\alpha)$  je  $\alpha$ -kvantil Studentovho rozdelenia pri  $n - 2$  stupňoch voľnosti. Interval spoľahlivosti na hladine  $1 - \alpha$  pre parameter  $\beta_1$  má tvar

$$\left( \hat{\beta}_1 - t_{n-2}(\alpha)s/\sqrt{\sum_{i=1}^n X_i^2 - n\bar{X}^2}, \hat{\beta}_1 + t_{n-2}(\alpha)s/\sqrt{\sum_{i=1}^n X_i^2 - n\bar{X}^2} \right).$$

Pri použití lineárnej regresie na naše dáta si za nezávislé premenné  $X_i$  zoberieme cenu jednotlivých sušienok študovaného segmentu D. Za závislé premenné  $Y_i$  zoberieme hmotnosť predaných sušienok v gramoch na jedného obyvateľa Českej republiky, teda pracujeme s rovnakými údajmi ako v celej kapitole. V nasledujúcej tabuľke 3.3 sú uvedené konkrétne hodnoty premenných.

Tabuľka 3.3					
$X_i$	$Y_i$	$X_i$	$Y_i$	$X_i$	$Y_i$
4.16	237.64	10.64	128.12	17.00	129.44
4.72	616.93	10.95	115.40	17.05	227.82
5.09	20.14	11.45	75.18	17.11	145.62
6.24	590.08	12.33	282.31	19.72	40.57
7.21	127.21	12.41	468.17	21.51	34.16
8.62	238.73	12.82	30.19	21.77	7.84
8.89	102.08	12.84	29.13	23.02	23.68
9.00	31.77	13.13	43.20	30.97	7.30
9.17	10.41	13.37	34.21	31.25	4.49
9.55	207.43	13.45	40.81	33.65	8.58
9.80	86.17	13.63	214.77	36.56	2.35
10.25	58.36	14.42	37.39	41.46	89.45
10.36	134.48	14.45	342.84	56.41	6.14

Predpokladáme, že platí model lineárnej regresie, a teda vyššie uvedené predpoklady sú splnené. Chceme otestovať hypotézu  $H_0 : \beta_1 = 0$ , t.j. že cena výrobku neovplyvňuje jeho predané množstvo.



Odhady parametrov dostaneme postupne pomocnými výpočtami:

$$n = 39,$$

$$\sum_{i=1}^{39} X_i = 636.43, \quad \sum_{i=1}^{39} X_i^2 = 405043.145,$$

$$\sum_{i=1}^{39} Y_i = 5030.599, \quad \sum_{i=1}^{39} Y_i^2 = 25306923.09,$$

$$\sum_{i=1}^{39} X_i Y_i = 5667.029, \quad \bar{X} = 16.319,$$

$$\hat{\beta}_0 = \frac{405043.145 \times 5030.599 - 636.43 \times 5667.029}{39 \times 405043.145 - 636.43^2} = 132.1498,$$

$$\hat{\beta}_1 = \frac{39 \times 5667.029 - 636.43 \times 5030.599}{39 \times 405043.145 - 636.43^2} = -0.1936,$$

$$s^2 = \frac{1}{37} (25306923.09 - 5030.599\hat{\beta}_0 - 5667.029\hat{\beta}_1) = 666033.18.$$

Pre  $\alpha = 0.05$  dostávame

$$|\hat{\beta}_1| \sqrt{\sum_{i=1}^n X_i^2 - n\bar{X}^2/s} = 0.149 < t_{37}(\alpha) = 2.021.$$

To znamená, že hypotézu  $H_0$  nemôžeme zamietnuť, respektívne nemôžeme s istotou tvrdiť, že predaj sušienok závisí na ich cene. Toto vyplýva aj z intervalu spoľahlivosti pre  $\beta_1$  na hladine 0.95, kde  $0 \in (-2.819, 2.432)$  a teda hypotézu  $H_0$  opäť nezamietame.

Vzhľadom k tomu, že test lineárnou regresiou nepotvrdil závislosť predaja sušienok na ich cene, ako sme predpokladali, tak sa ešte bližšie pozrieme na rozdelenie reziduí, to by malo byť normálne. Reziduum, označíme ho  $u_i$ , je rozdiel medzi pozorovanou hodnotou  $Y_i$  z pôvodných dát a predikovanou hodnotou  $y_i$  vypočítanou pomocou

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i.$$

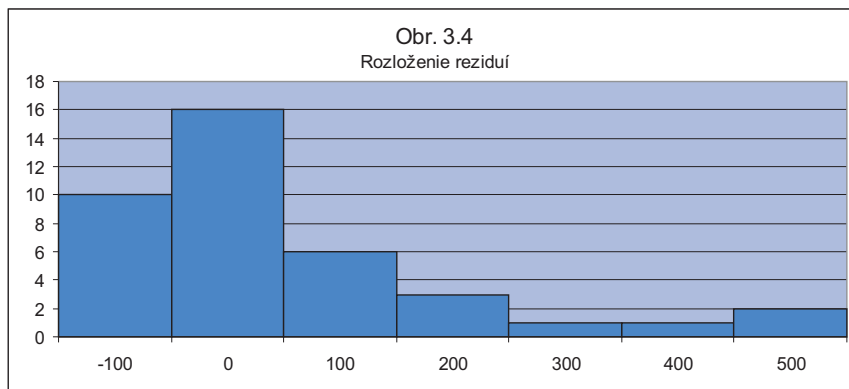
Hodnoty reziduí teda dostaneme zo vzťahu

$$u_i = Y_i - y_i$$

a v tabuľke 3.4 sú už uvedené vypočítané reziduá pre naše dáta.

Tabuľka 3.4								
$Y_i$	$y_i$	$u_i$	$Y_i$	$y_i$	$u_i$	$Y_i$	$y_i$	$u_i$
237.64	131.34	106.30	128.12	130.09	-1.97	129.44	128.86	0.58
616.93	131.24	485.69	115.40	130.03	-14.63	227.82	128.85	98.97
20.14	131.16	-111.02	75.18	129.93	-54.75	145.62	128.84	16.79
590.08	130.94	459.13	282.31	129.76	152.55	40.57	128.33	-87.76
127.21	130.75	-3.55	468.17	129.75	338.42	34.16	127.99	-93.83
238.73	130.48	108.25	30.19	129.67	-99.47	7.84	127.94	-120.09
102.08	130.43	-28.35	29.13	129.66	-100.53	23.68	127.69	-104.01
31.77	130.41	-98.64	43.20	129.61	-86.41	7.30	126.15	-118.86
10.41	130.37	-119.96	34.21	129.56	-95.35	4.49	126.10	-121.61
207.43	130.30	77.13	40.81	129.55	-88.73	8.58	125.64	-117.05
86.17	130.25	-44.08	214.77	129.51	85.26	2.35	125.07	-122.72
58.36	130.17	-71.80	37.39	129.36	-91.97	89.45	124.12	-34.67
134.48	130.14	4.33	342.84	129.35	213.49	6.14	121.23	-115.09

Pre zistenie normality reziduí je vhodné vykresliť histogram ich rozloženia, ktorý môžeme vidieť na ďalšom obrázku 3.4. Z histogramu je zrejmé, že rozloženie reziduí okolo nuly nie je rovnomerné, ale má odchylku. Bolo by možné urobiť vhodnú transformáciu vektorov premenných  $\mathbf{X}$  alebo  $\mathbf{Y}$ , prípadne oboch a ešte raz hypotézu otestovať. My však test opakovať nebudeme, ale zoberieme do úvahy možné ovplyvnenia odchylky pri počítaní regresných odhadov a ich testovaní.



Testovaním závislosti medzi cenami sušienok a ich predajom pre všetky kraje sme dospeli k rovnakým výsledkom. Testy hypotéz nepotvrdili, že cena výrobku výrazne

ovplyvňuje jeho odbyt. Toto zistenie sa môže zdať prekvapivé, pretože z pohľadu na grafy ukazujúce hmotnosť predaných sušienok v závislosti na cene v Prílohe 2 je vidieť, že s rastúcou cenou trend predaja sušienok v každom kraji z globálneho hľadiska klesá. Na druhej strane, výrobky, s ktorými sa obchoduje na rýchloobrátkových trhoch, teda aj sušienky, sa pohybujú v nízkych cenových hladinách a bez väčších rozdielov v cene. To znamená, že kupujúci, stále hovoríme o spotrebiteľoch nakupujúcich v medzinárodných obchodných reťazcoch, necítia potrebu na tomto tovare šetriť, ale vyberajú si sebou osvedčené a tradičné výrobky.

# Kapitola 4

## Záver

Situácia na trhu rýchloobrátkového tovaru sa veľmi rýchlo mení, čo výrobcov núti neustále prehodnocovať cenovú stratégiu alebo plán výroby, a preto sú prieskumy a analýzy v tomto odvetví veľmi dôležité. Cieľom našej práce bolo urobiť analýzu z oblasti rýchloobrátkového trhu sušienok a oblátiek. Skúmali sme ako sa mení predaj výrobkov v čase, či už v priebehu sledovaných štyroch rokov alebo sezónne a tiež sme sa zaoberali závislosťou predanosti výrobkov na mieste ich odbytu.

Začali sme analýzou celkovej predanej hmotnosti sušienok jednotlivých segmentov v závislosti na čase. Vyrovnaním časových radov centrovanými kľúčovými priermi sme získali krivky popisujúce trend predaja pre všetky segmenty sušienok. Ako sme predpokladali, väčšina segmentov vykazovala rastúci trend. Výnimkou boli iba výrobky skupiny B, kde mal trend miernu klesajúcu tendenciu. Tento výsledok ale pre nás nebol prekvapivý, pretože sa jednalo o sušienky obľúbené v úzkej skupine spotrebiteľov.

Ďalej sme sa zaoberali sezónnosťou predaja, teda študovaním sezónnej zložky časových radov a výsledky opäť vo veľkej časti odpovedali tomu, čo sme očakávali. Nárast predaja bol v jednotlivých segmentoch individuálny. Predaj väčšiny sušienok poklesol v decembri, pretože v okolí Vianoc si domácnosti pečú vlastné sladkosti. Nárast predaja tu dosiahol iba segment F, ktorý obsahoval výrobky používané pri výrobe domácich sladkostí a koláčov.

Prekvapivejšie výsledky vyšli pri priestorovej analýze, kde sme opäť skúmali celkovú predanú hmotnosť sušienok, tentokrát iba pre segment D a vo vzájomnej závislosti na krajoch, priemerných mzdách v krajoch a cenách jednotlivých sušienok. Po rozdelení výrobkov do troch cenových skupín sme porovnali rozdiely medzi pomermi predanej hmotnosti cenovej skupiny k priemeru hmotností predaných sušienok danej

skupiny za všetky kraje Českej republiky okrem Prahy, ktorá dosahovala voči ostatným krajom príliš vysoké hodnoty. Ďalším porovnaním s priemernou mzdou sme dospeli k záveru, že životná úroveň kraja sa na predaji sušienok výrazne neodráža.

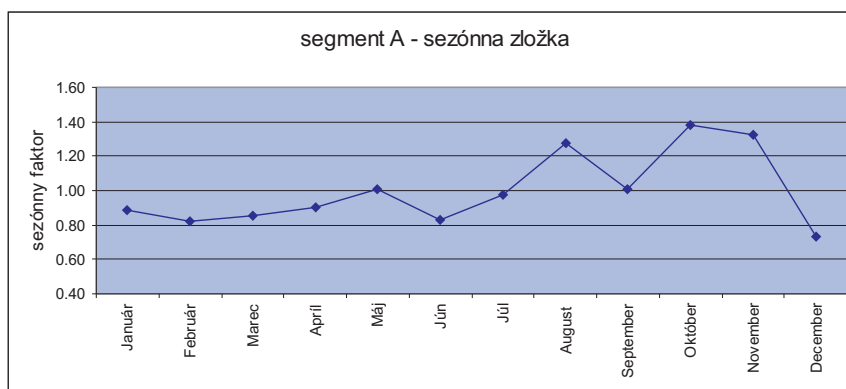
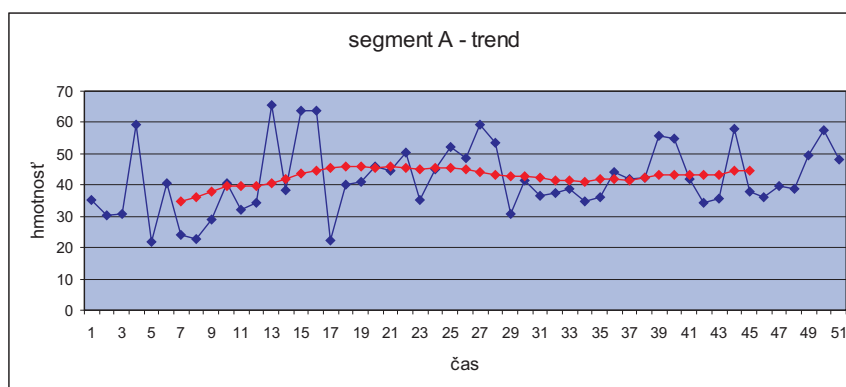
Analýzu dodaných dát sme ukončili lineárnou regresiou, testovaním, či množstvo predaných sušienok závisí na ich cene. Avšak, kvôli nesplneným predpokladom modelu, nemôžeme výsledky testovania hypotézy brať v úvahu.

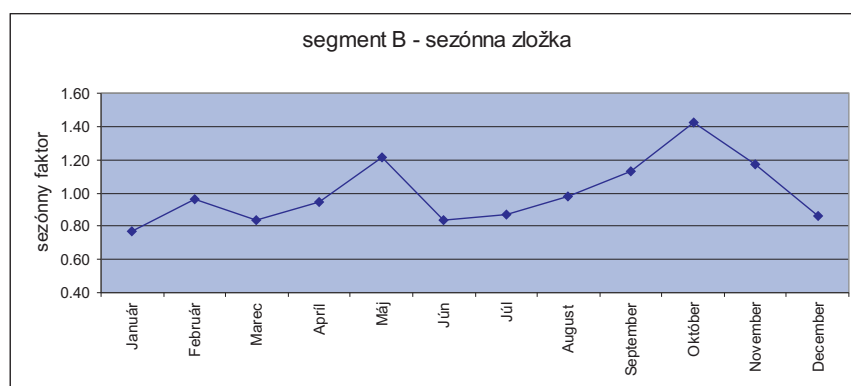
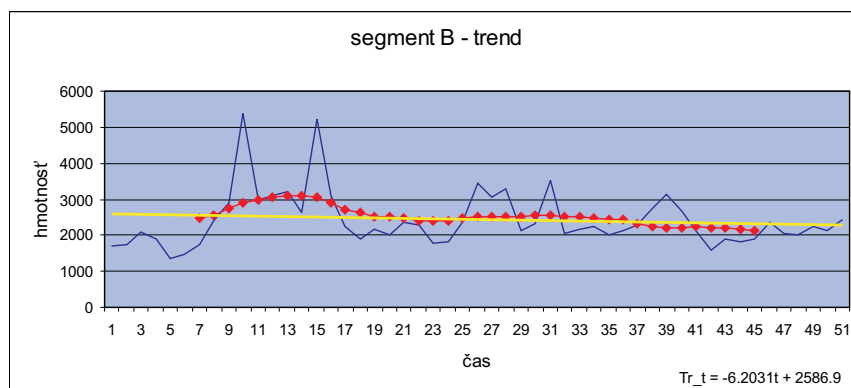
Existujú ešte ďalšie spôsoby analýzy trhu rýchloobrátkového tovaru a iné závislosti, ktoré sa dajú na ňom skúmať, ale ku ktorým sme sa už nedostali. Napríklad by bolo možné podrobnejšie sa venovať priestorovej závislosti a zohľadniť vzájomnú interakciu krajov. Tiež by sa dali skúmať zmeny a vývoj priestorovej závislosti predaja v jednotlivých krajoch v čase. Jedným z najlepších vylepšení analýzy by však bolo pracovať s rozsiahlejším počtom dát, ktoré by pokrývali celý trh, nielen s dátami z medzinárodných obchodných reťazcov, ktoré sme mali k dispozícii.

# Literatúra

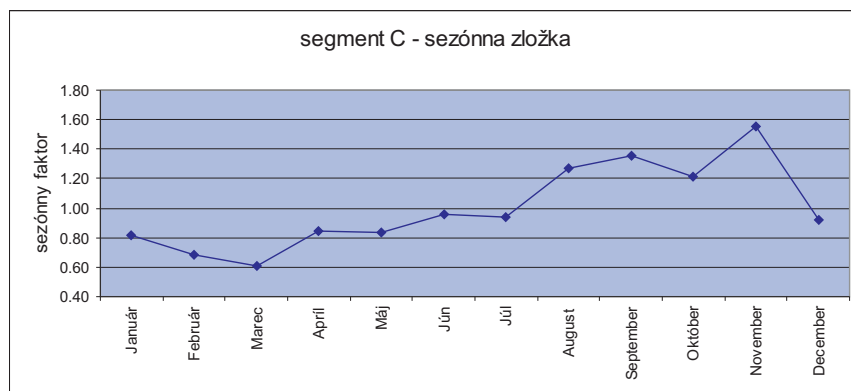
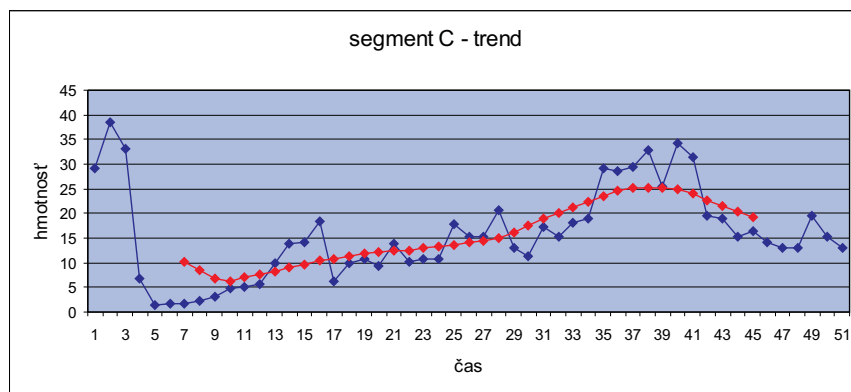
- [1] Anděl, J. (2007): *Základy matematické statistiky*, MATFYZPRESS, Praha, ISBN 80-7378-001-1.
- [2] Cipra, T. (1986): *Analýza časových řad s aplikacemi v ekonomii*, SNTL, Praha a ALFA, Bratislava.

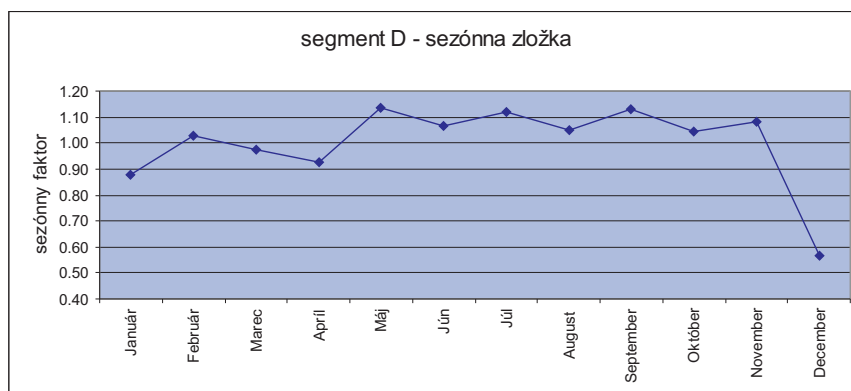
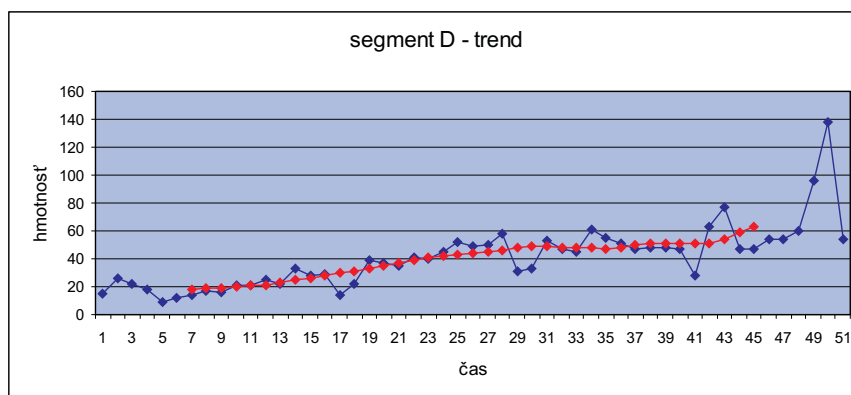
# Príloha 1

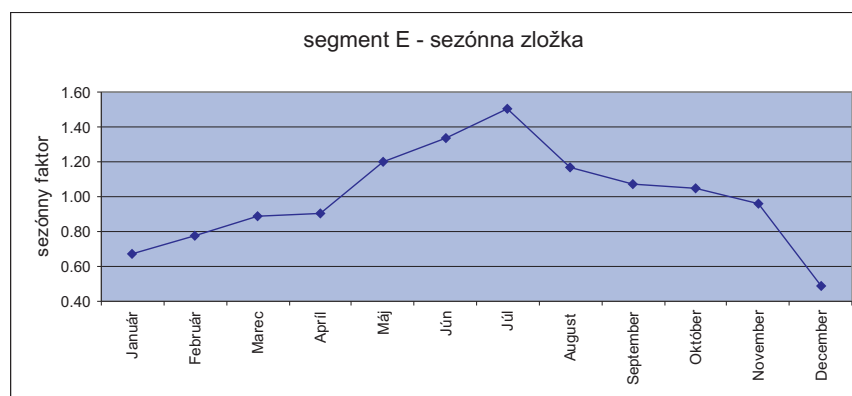
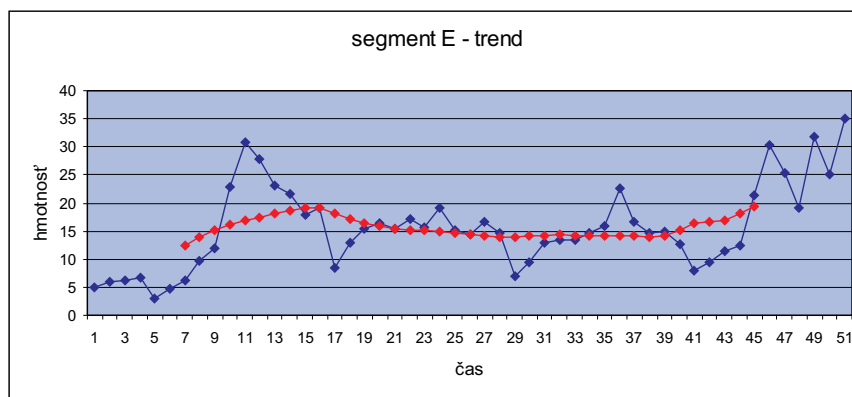


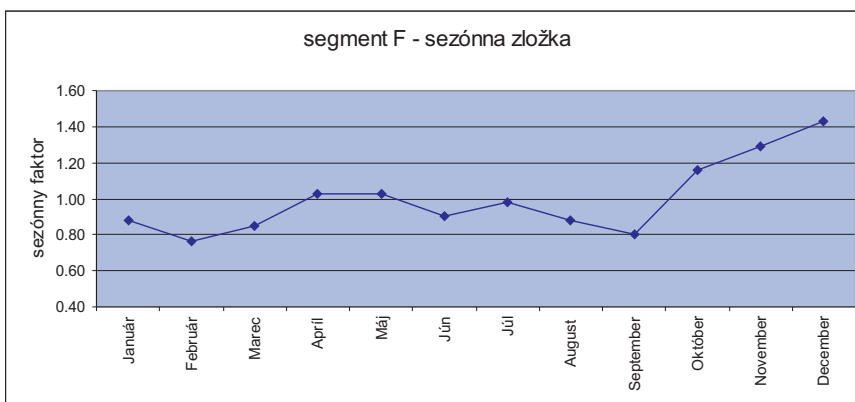
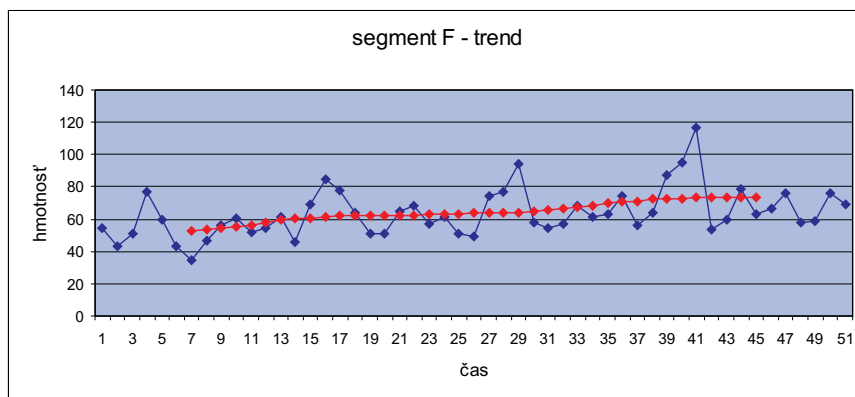


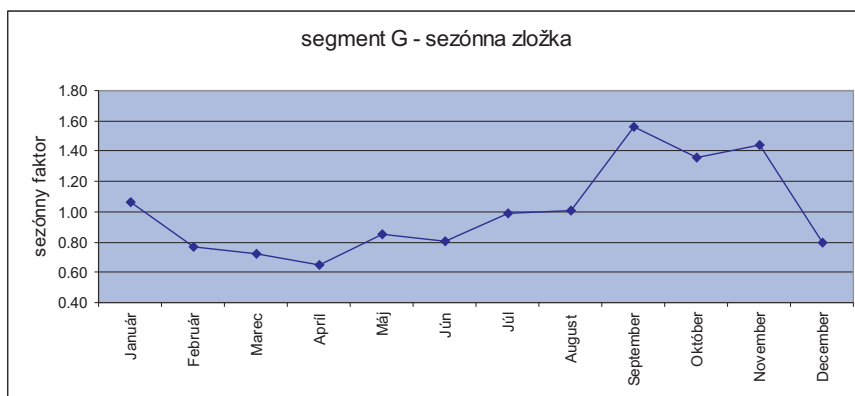
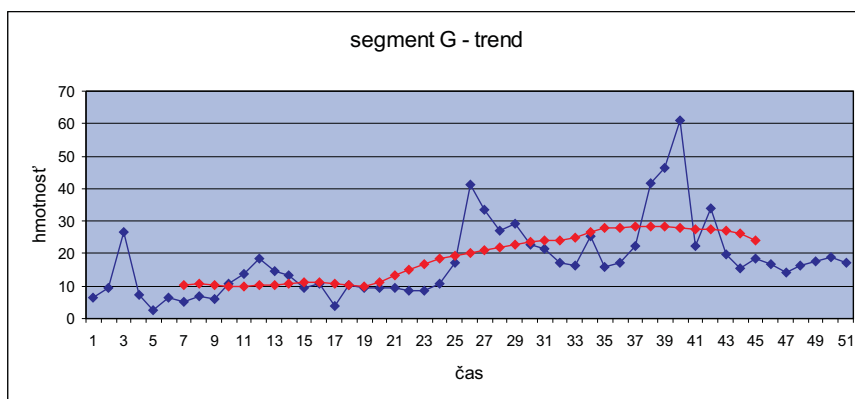


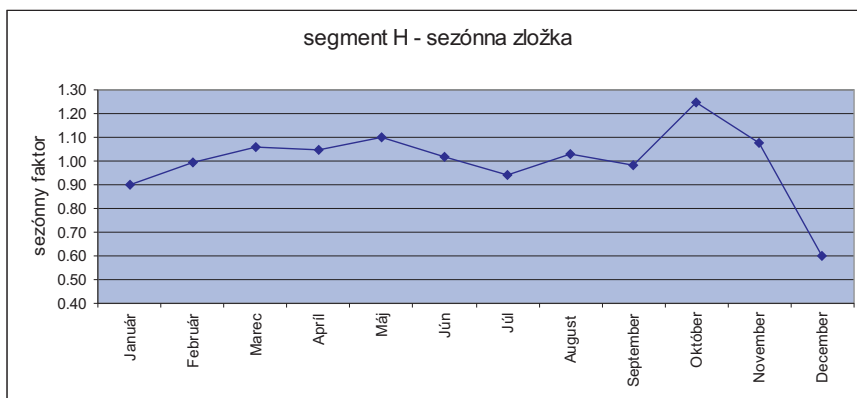
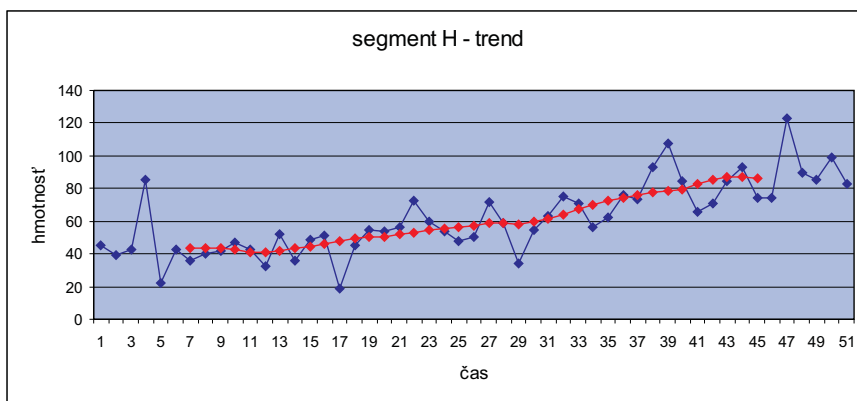


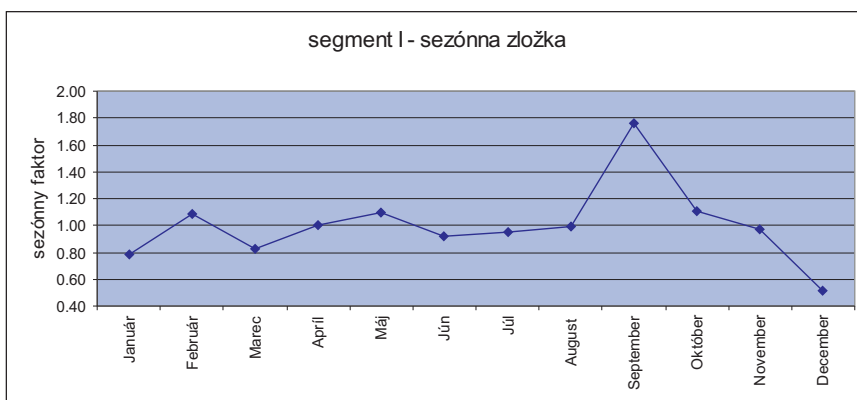
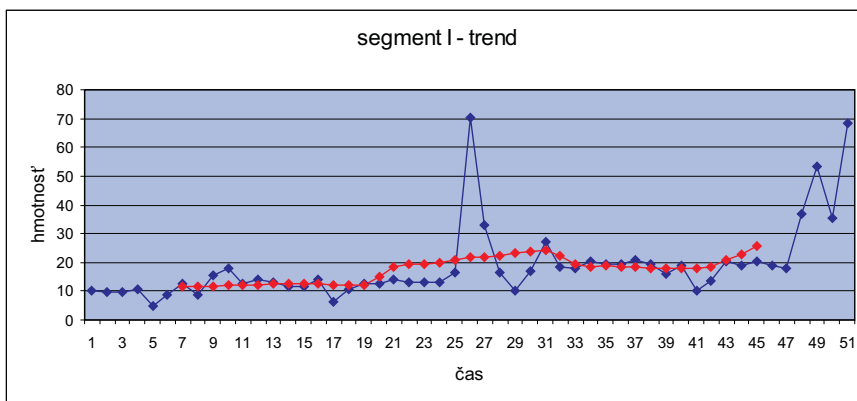


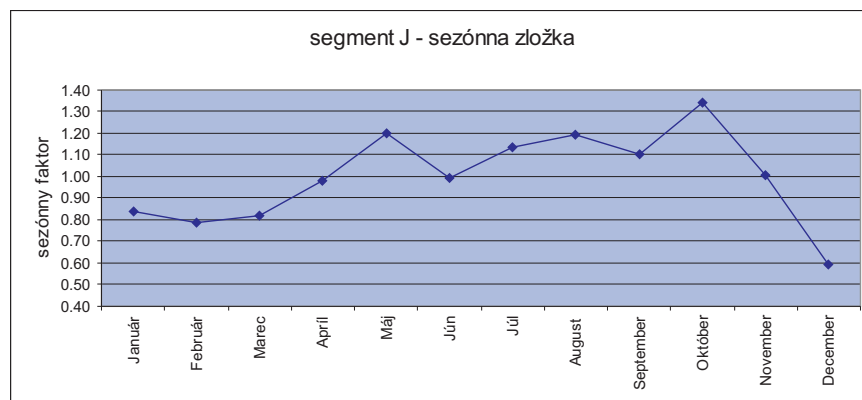
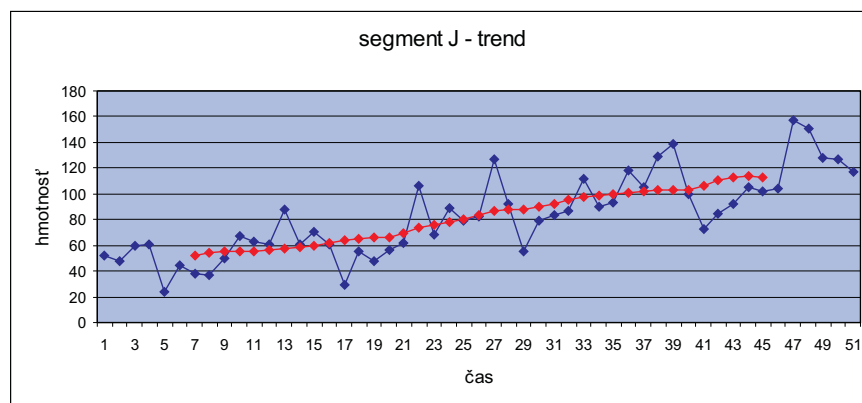




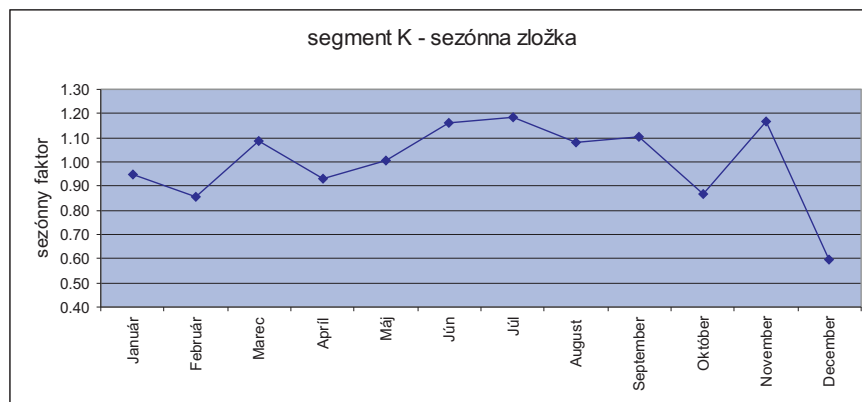
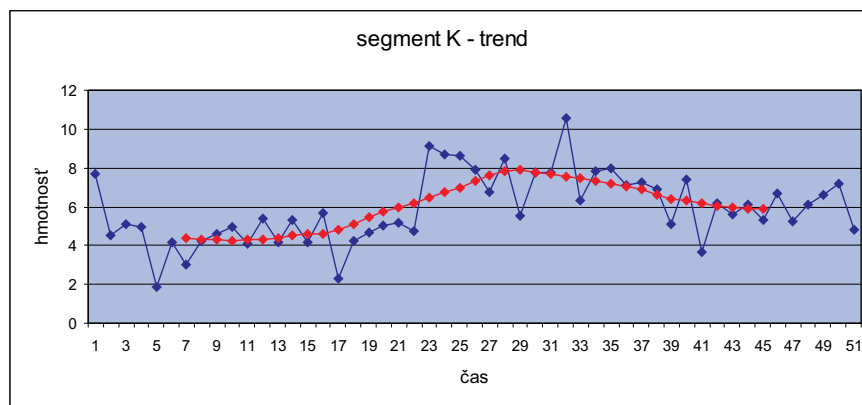












## Príloha 2

Segment D			
Ean kód	cena/100g	Ean kód	cena/100g
Ean545	4.16 (L)	Ean102	13.13 (P)
Ean541	4.72 (L)	Ean910	13.37 (P)
Ean424	5.09 (L)	Ean103	13.45 (P)
Ean631	6.24 (L)	Ean544	13.63 (P)
Ean546	7.21 (L)	Ean85	14.42 (P)
Ean730	8.62 (L)	Ean657	14.45 (P)
Ean162	8.89 (L)	Ean644	17.00 (P)
Ean919	9.00 (L)	Ean630	17.05 (P)
Ean423	9.17 (L)	Ean643	17.11 (P)
Ean542	9.55 (L)	Ean155	19.72 (P)
Ean632	9.80 (L)	Ean909	21.51 (D)
Ean537	10.25 (P)	Ean105	21.77 (D)
Ean642	10.36 (P)	Ean104	23.02 (D)
Ean543	10.64 (P)	Ean51	30.97 (D)
Ean461	10.95 (P)	Ean53	31.25 (D)
Ean154	11.45 (P)	Ean80	33.65 (D)
Ean648	12.33 (P)	Ean945	36.56 (D)
Ean645	12.41 (P)	Ean646	41.46 (D)
Ean15	12.82 (P)	Ean167	56.41 (D)
Ean14	12.84 (P)		

